

Received March 1, 2021, accepted March 16, 2021, date of publication March 22, 2021, date of current version March 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3067677

# Classification of Germination Images of Pear Pollen Using Random Forest and Convolution Neural Network Models

UNG YANG<sup>ID1</sup>, SEUNGWON OH<sup>ID2</sup>, SEUNG GON WI<sup>ID1</sup>, BOK-RYE LEE<sup>ID1</sup>,  
SANG-HYUN LEE<sup>ID1,3</sup>, AND MIN-SOO KIM<sup>ID2</sup>

<sup>1</sup>Asian Pear Research Institute, Chonnam National University, Gwangju 61186, South Korea

<sup>2</sup>Department of Mathematics and Statistics, Chonnam National University, Gwangju 61186, Republic of Korea

<sup>3</sup>Department of Horticulture, College of Agriculture and Life Sciences, Chonnam National University, Gwangju 61186, South Korea

(Corresponding authors: Sang-Hyun Lee (pear@chonnam.ac.kr) and Min-Soo Kim (kimms@chonnam.ac.kr)

(Ung Yang and Seungwon Oh are co-first authors.)

The present study was supported by a grant from the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries (IPET) through Export Promotion Technology Development Program (No. 617075-5) and Advanced Production Technology Development Program (No. 318065-03) funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA).

**ABSTRACT** Verifying pollen germination using microscopic images is a difficult task. It is usually time-consuming and may entail reduced accuracy and reproducibility. Therefore, in this study, we used random forest (RF) and convolutional neural network (CNN) models to perform image classification on raw data corresponding to pollens with different germination rates; the data were obtained via flow cytometry. A heat map, which was based on the RF analysis results, showed that the variables that significantly influenced the classification decision between NG and 60G categories were mainly located in the center and top-right regions of the  $30 \times 30$  pixel image. Additionally, a variable importance plot showed that among the 900 input variables, pixel\_316 was the variable that contributed the most toward prediction. Gradient-weighted class activation mapping was used to visualize the class activation maps of the CNN model. The bottom-left region of the activation map was activated in the NG image. However, the 60G image showed that not only the bottom-left region but also the top-right region was activated. Both the models classified the input images into NG and 60G categories with high accuracy. However, considering that the RF model does not reflect the characteristics of adjacent variables, the CNN model is more appropriate for classifying pollen germination images corresponding to pollen with various germination rates into distinct classes. Taken together, these results suggest that the CNN model can provide a reliable method for verifying the pollen performance.

**INDEX TERMS** Classification, CNN, image, pollen, RF, variables.

## I. INTRODUCTION

Most pear cultivars require insect-mediated pollination, but insect populations are in rapid decline caused by recent environmental changes such as habitat loss, environmental pollution and climate change [1]–[5]. Thus, the artificial pollination has been compelled to offset the decline in insect pollinators and ensure satisfactory crop yields in many commercial orchards [6], [7]. It is necessary to find an easy and reliable method to assess viability of stored pollen before applying it to artificial pollination, as pollen viability, which

The associate editor coordinating the review of this manuscript and approving it for publication was Ioannis Schizas<sup>ID</sup>.

affects pollination efficiency, may be reduced during storage period.

Numerous methods have been developed to assess pollen viability, of which in vitro germination is the most commonly used method [8]. This method is easy and simple, but it is necessary to determine the optimal conditions for pollen germination. Furthermore, there is a limitation in that the germination rate in vitro is determined using only the pollens identified from microscopic image, not all pollen used in the assay [9]. In this method, the germination rate is calculated as the proportion of pollen grains germinated to the total number of pollen grains existed in microscopic image.

In a previous study [9], we proposed a method of evaluating pollen viability through flow cytometry in order to reduce the drawbacks of conventional method (i.e., *in vitro* germination) in respect of reliability, subjectivity of counting, and analysis time, and to improve the accuracy of measurement. The presence or absence of the pollen tubes (i.e., pollen has germinated or not) can be determined by flow cytometry, which can distinguish individual pollen according to their size (FSC parameter) or internal complexity (SSC parameter). This became evident through our previous result, which confirmed the difference in density distribution on the dot plot between pollen samples with different viability. In addition, since 50,000 pollen grains are used per flow cytometric analysis, reproducibility and accuracy can be ensured, unlike conventional method of measuring germination rate from 200-300 pollen grains. Coordinate values of the FSC (x-axis) and SSC (y-axis) parameters derived from 50,000 pollen grains following flow cytometric analysis are transformed into a germination image, which is suitable for fitting a neural network classification, through data refining process. It may therefore provide a new perspective to establish a reliable method of verifying pollen performance, as germination images of pear pollen, which can vary depending on the germination rate.

Unlike structured data, the image data are high-dimensional. So, the problem may occur with the number of parameter being increased in conventional machine-learning model, when dealing with the image data. For that reason, shallow neural networks using fully connected layers are not suitable for classifying the image data.

The most used model in machine learning is random forests (RF), which use a random subspace method based on decision trees [10], [11]; this concept entailed a method of building a forest of uncorrelated trees using a CART-like procedure, combined with randomized node optimization and bagging. RFs offer many advantages, namely, satisfactory predictive performance, robustness to both noise and overfitting, and extraction of information regarding the variables that are important for the classification [11]–[13].

Generally, in image data, adjacent pixels are spatially correlated [14], and loss of information may occur during vectorization because of the spatial characteristics of pixels. However, conventional machine-learning techniques do not reflect the spatial characteristics of the image and rather use each pixel as a variable, which in turn cause increased the time and number of parameters required for learning [15]. Therefore, LeCun *et al.* [16] proposed a convolutional neural network (CNN) that reflected the spatial information of the image and overcame the limitations of the existing methods. Unlike other fully connected neural network models, CNNs can effectively recognize the relationship between pixels while maintaining the spatial information of the input image. In the ImageNet Large Scale Visual Recognition Challenge, Alexnet [17], which used a CNN-based architecture, was ranked the first in 2012, showing exemplary performance compared with conventional computer-vision techniques.

This effectiveness is achieved by using a convolution layer with multiple filters and intensifying the extracted features after convolution layer stacked using a pooling layer [17]. The filter parameters used for feature extraction are shared for all the input data; thus, fewer parameters are required and the learning time is shortened, as compared with the number of parameters and learning time of a regular neural network [18]. In a CNN model, the convolutional layer and pooling layer used to extract features are freely configurable. Finally, the extracted features are fed to a fully connected layer, following which the input data are classified.

Therefore, one can use machine-learning techniques such as RF and CNN that perform excellent image recognition and classification, to develop a novel validation system that can discriminate the pollen performance. This approach can be achieved by training both the models with germination images, which extracted from pollens with different germination rate, and then performing classification on germination images into distinct classes according to germination rates.

Over the past few years, several studies using machine-learning algorithms have been conducted to monitoring of the airborne pollen or to develop automatic classification system of pollen grains [19]–[21]. There are also many studies that have utilized CNN models to classify various pollen species [19]–[24]. However, our research is focused on establishing a reliable method for assessing pollen viability using germination images rather than identifying the pollen species covered in the palynology fields.

The purpose of this study is to determine whether the input images are classified into two categories of NG and 60G by RF and CNN models that have learned germination images extracted from two pollen groups with different germination rates. To achieve the goal, 1) the raw input data were transformed into matrix and vector formats for machine learning. 2) The transformed data was trained in RF and CNN models. 3) Classification was performed on the input images into two class labels through the classifier of both models trained. 4) The extracted features were visualized to ensure which part(s) of the pollen germination image were considered in the final classification decision in each model. Thus, the results obtained here may provide a framework for establishing a reliable method to assess pollen performance from germination images. Classification on pollen germination images into multiple labels according to germination rates can be achieved through a classifier that has learned features reflecting the difference in germination images.

## II. MATERIALS AND METHODS

### A. POLLEN COLLECTION

Unopened, balloon-stage flowers were collected in April 2019 (5 days before full bloom) from the pear trees of the “Wonwhang” cultivar, which is grown in commercial orchards (Gongsan, Naju, Korea). The anthers from the pear flowers that were not completely open were harvested. Subsequently, to dry and release their pollen within 24 to 36 h, we laid

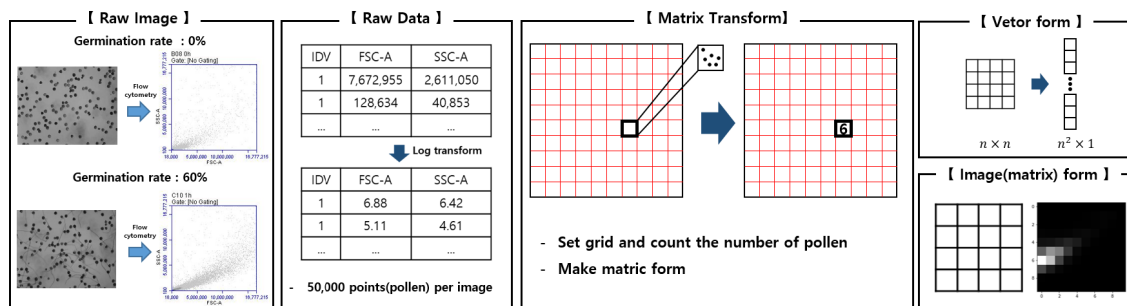


FIGURE 1. Refining process for classification of pollen germination image.

out the anthers on a black kentpaper in the anther dehiscence room at 20 °C with the relative humidity of 50%. When more than 80% of the anthers were dehiscent, they were transferred from the kentpaper to a stainless-steel bowl equipped with a 100-mesh sieve (aperture: 0.149 mm). Acetone was then poured into the bowl, and the pollens were gently sifted through the 100-mesh sieve. Subsequently, the acetone supernatant was carefully discarded from the bowl, and the residual solvent was volatilized. The pollen samples, each weighing 10 g, were placed in tightly sealed containers and frozen at -60 °C, until further use.

**B. FLOW CYTOMETRY**

Unless otherwise stated, an assay was conducted by removing the pollen container from a deep-freezer and thawing the pollen overnight at 4 °C in a desiccator. Subsequently, pollen at a concentration of 2.5 mg·mL<sup>-1</sup> was suspended in a liquid medium that contained 10% sucrose (w/v), 0.4 mM boric acid, and 1 mM calcium nitrate, and the mixture was then incubated at 25 °C for 3 h. Data were acquired by analyzing 1 mL of the pollen culture per sample every hour by using the Accuri C6 flow cytometer (Becton Dickinson, USA), according to the instructions provided by the manufacturer. The Accuri C6 flow cytometer was generously provided for this study by Damyang Agricultural Technology Center. The germinated pollens were plotted using the FSC and SSC parameters and visualized as dots on a dot plot by analyzing 50,000 pollen grains per hour. The CSV data of the germinated pollen obtained here were used as raw data for performing machine-learning analysis; they contained the coordinate values of the FSC and SSC parameters for 50,000 pollen grains.

**C. DATA REFINING**

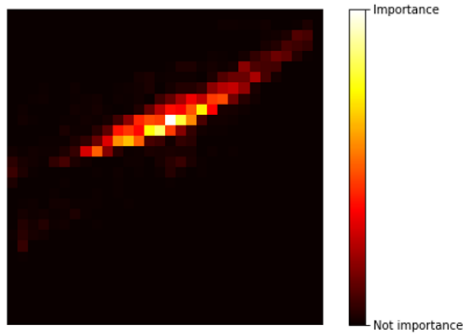
To reduce the range of data, the raw data were log-transformed prior to applying them to the classification model. A grid was established to convert the CSV data, which contained the coordinates of the FSC and SSC parameters, into matrix form. Notably, the spatial features of an image are reflected increasingly well with an increase in the grid size. However, larger grid sizes often result in increased processing

time. Thus, we loaded the CSV data within grids with a size of 30 × 30 pixels. This is because the grid size of 30 × 30 is sufficient to represent the density distribution on the dot plots reflecting whether the pollen has germinated. The number of pollen grains that belong to the set grid range was computed and then divided by the maximum value in the matrix to normalize all the pixel values to lie between 0 and 1 before using the pixel values as the value of the corresponding row and column. The matrix-form data were either used to train the CNN model to extract the spatial features of the image, or applied to the RF model after converting the data into vector form to identify the influence of pixels on the classification decision. The process of image transformations required for machine learning is illustrated in Fig. 1.

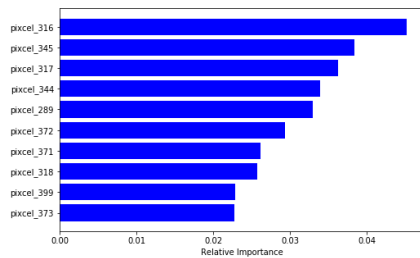
The entire data set was randomly divided into training and testing sets in the ratio of 70:30. The images of the pollen samples analyzed on the dot plot immediately after suspension in the germination medium were used as image data of non-germinated pollen (NG). Additionally, the images of the samples analyzed after incubation for 1 h were used as image data of pollen that had germinated to the extent of over 60% (60G). The average germination rate of the pollen samples used in this study was calculated as approximately 60% after the incubation for 1 h (data not shown). For the training set, 168 images for 60G and 104 for NG were used. For the testing set, 72 images for 60G and 45 for NG were used.

**III. RESULTS AND DISCUSSION**

To fit the RF model, the input images were vectorized with the size of 30 × 30 pixels and the model was fitted using 272 images and 900 pixel variables. The optimal parameters were obtained via hyperparameter tuning with three-fold cross validation. The final parameters used in the RF model are as follows: bootstrap = True; n-estimators (i.e., the number of trees used) is 1,000; the maximum number of leaf nodes is 64; criterion is entropy. The heat map in Fig. 2a indicates the relative importance of individual pixels that affect the classification decision. The variables that significantly affect the classification between NG and 60G were mostly located in the center and top-right regions of the images.



(a) Heat map illustrates the relative importance of pixel variables.



(b) The top 10 pixels that contributed the most to the prediction among the input variables applied to the RF model.

**FIGURE 2. Variable importance plot (VIP).**

A variable importance plot (VIP) shows the top 10 pixels that contributed the most to the prediction among the input variables applied to the RF model (see Fig. 2b). From the plot, pixel\_316 seems to be the most significant variable, followed by pixel\_345, pixel\_317, pixel\_344, and so on. These results show the information that the RF model has learned to distinguish NG and 60G, and they may provide additional insight into the major contributing pixels that indicate whether the pollen has germinated.

The CNN architecture used and its training process are illustrated in Fig. 3. It comprises convolution layers to extract features and a global average pooling (GAP) layer to classify the input data into two distinct classes, i.e., NG and 60G. The input data were transformed into the output after stacking several convolution and pooling layers. For CNN analysis, 272 images (i.e., 168 images for 60G and 104 for NG), which were transformed into matrix form, were processed through 3 convolution layers. The features maps, which ranged from low to high level, were extracted upon passing the input through each convolutional layer. Generally, high-level features extracted from the last convolution layer were used for the classification task after transferring them to the fully connected layer. The feature maps generated from each convolution layer were represented in an increasingly abstract manner as they passed through each layer. Therefore, it is typically difficult to know which part(s) of the input image influenced the final classification decision.

Additionally, it remains difficult to explain the predictions of the CNN owing to the lack of interpretability. Therefore, the visualization of the activated features is necessary to verify that the resulting classifier makes decisions on the basis of appropriate features present in the training data, although the CNN makes accurate predictions. Gradient-weighted class activation mapping (Grad-CAM) was used to visualize the effect of each input image on the classification decision [25]. For feature visualization, the gradient and output information of the last convolution layer was used. The heat map is a class activation map. It indicates the importance of each region of the image with regard to the classification decision by highlighting the image regions responsible for a particular prediction, as illustrated in Fig. 4. The bottom-left region of the activation map was activated in the image that corresponded to NG. However, in the image that corresponded to 60G, not only the bottom-left region but also the top-right region was activated. This helps understand the manner in which the network produced the output more intuitively because class prediction mainly depends on the activation of the last convolution layer.

For both the models used in this study, almost all of the pollen germination images in the testing set were classified into NG and 60G with high accuracy by using the above-mentioned parameters and classification rules. The RF model yielded a classification accuracy of 100.0% for 272 training images, and the CNN model obtained a classification accuracy of 99.6% owing to one misclassification. For 117 testing images, the RF model achieved the classification accuracy of 99.1% except for one misclassification, whereas the CNN model exhibited the classification accuracy of 100.0% for all the samples (Table 1, 2).

This indicates that both the models could appropriately classify the input data into distinct classes. Additionally, both the models may be used to learn multiple categories to classify the pollen germination images with various germination rates in a future study. By visualizing the activated features, one overcomes the limitation of CNNs that they function as black-box models whose results are difficult to interpret [26]. Additionally, one can know what the model has learned to classify an image. The RF model performs learning by converting input variables into the vector form. The RF analysis revealed that the RF model classified the input data into NG and 60G categories with high accuracy, although the characteristics of adjacent pixel variables could not be reflected. In data processing using machine learning, the image data are susceptible to distortions such as rotation or movement. Additionally, when the number of samples or the number of classification categories increases, classification errors may occur in the RF model where uses pixels as a parameter, owing to the lack of spatial information. However, important CNN concepts, including sparse connectivity, parameter sharing, subsampling, and local receptive field, ensure that the CNN is not affected by the movement, scaling, and distortion of the input data [16], [17], [27].

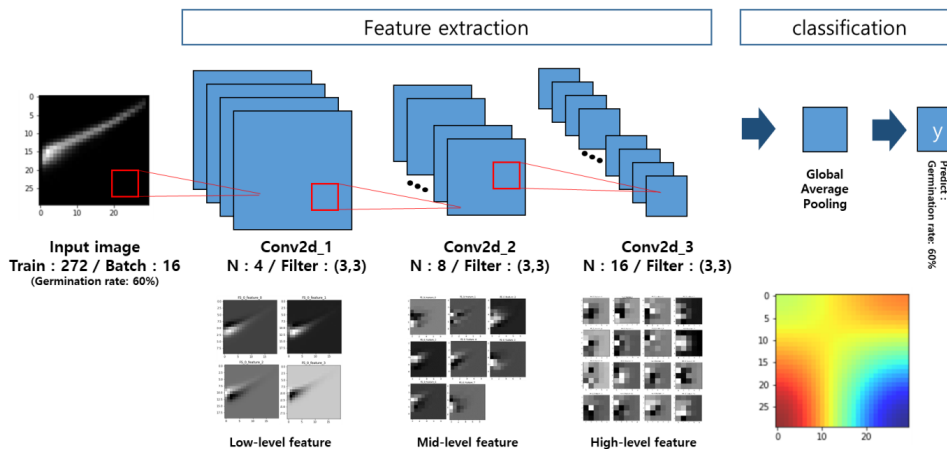
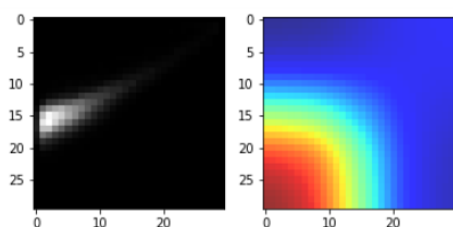
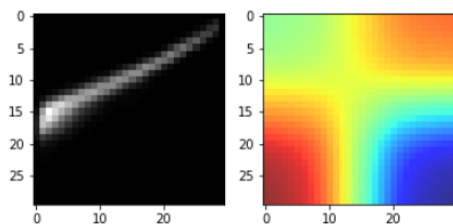


FIGURE 3. CNN model architecture.



(a) Germination rate 0%



(b) Germination rate 60%

FIGURE 4. Grad-CAM.

TABLE 1. CNN model classification result in test set.

	real	predict	
		0%	60%
	0%	72	0
	60%	0	45

TABLE 2. RF model classification result in test set.

	real	predict	
		0%	60%
	0%	72	0
	60%	1	44

The VIP provides the list of the most important variables in descending order with decrease in average entropy. The importance of variables is assessed on the basis of the part

TABLE 3. Accuracy of distorted images.

predict	True	D3	D6	U3	U6	R3	R6
CNN	60G	60G	60G	60G	60G	60G	60G
RF	60G	60G	NG	NG	NG	60G	60G

that represents the difference on each variable throughout the entire data set. The center and top-right regions of the pixel images, which show the difference between NG and 60G, significantly affected the classification (see Fig. 2a). This shows that the decision rules applied entirely to the input data, indicating that the algorithm used in the model was globally interpretable. Notably, global interpretability helps understand the entire relationship between the input and output; however, it may be approximate or based on average values [28]. Meanwhile, the result of Grad-CAM (see Fig. 4), which shows the features activated in each sample, has a meaning of local interpretation unlike VIP has shown.

The local interpretation focuses on the details in each image that might be overlooked in the global interpretation [29]. Thus, compared with the RF model, the CNN model which extracts features from each image and uses them for classification is considered more appropriate for classifying the pollen germination images with various germination rates into distinct classes.

Robustness is important because image data has sensitive characteristics even with small distortion. To compare the robustness, classification task was performed with two models following the germination image corresponding 60G was shifted up, down, and right by 3 and 6 spaces, respectively. The CNN model showed robust classification performance even in distorted images, whereas misclassification occurred in RF into NG in case of D6, U3, and U6 (Table 3). Therefore if there is distortion in input image, the performance of RF is not suitable for classification on pollen germination images because it does not reflect spatial information.

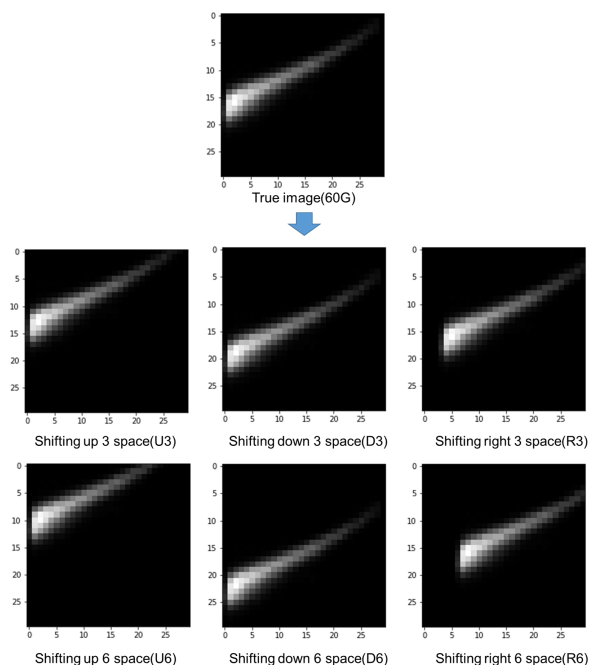


FIGURE 5. Distorted image using shifting.

Classification of pollen germination images using CNN model revealed to be suitable for establishing a reliable method to assess pollen performance. In this study, image classification was performed on raw data with difference in pollen germination rates by using both RF and CNN models. The results of VIP and Grad-CAM showed which region of the image affected the final classification decision in each model. Both RF and CNN models showed high classification accuracy in training and testing data. While the RF shows the general relationship between the input and the output learned by a model, the CNN focuses on how individual predictions are made by a model. Considering the spatial characteristics of the image, the CNN model is therefore more suitable for classifying pollen germination images with multiple categories into distinct classes in future study. However, for a detailed prediction of the germination rate, further studies are required to develop a classification and prediction model that reliably validates the pollen performance.

## REFERENCES

- [1] G. D. Powney, C. Carvell, M. Edwards, R. K. A. Morris, H. E. Roy, B. A. Woodcock, and N. J. B. Isaac, "Widespread losses of pollinating insects in Britain," *Nature Commun.*, vol. 10, no. 1, pp. 1–6, Dec. 2019.
- [2] P. Soroye, T. Newbold, and J. Kerr, "Climate change contributes to widespread declines among bumble bees across continents," *Science*, vol. 367, no. 6478, pp. 685–688, Feb. 2020.
- [3] S. G. Potts, J. C. Biesmeijer, C. Kremen, P. Neumann, O. Schweiger, and W. E. Kunin, "Global pollinator declines: Trends, impacts and drivers," *Trends Ecol. Evol.*, vol. 25, no. 6, pp. 345–353, Jun. 2010.
- [4] M. Y. Benoit Geslin, M. A. Aizen, N. Garcia, A. J. Pereira, B. E. E. Vaissiere, and L. A. Garibaldi, "The impact of honey bee colony quality on crop yield and farmers' profit in apples and pears," vol. 248, pp. 153–161, Oct. 2017.
- [5] M. T. Fountain, Z. Mateos-Fierro, B. Shaw, P. Brain, and A. Delgado, "Insect pollinators of conference pear (*Pyrus communis* L.) and their contribution to fruit quality," *J. Pollination Ecol.*, vol. 25, no. 10, pp. 103–114, 2019.
- [6] V. Pinillos and J. Cuevas, "Artificial pollination in tree crop production," in *Horticultural Reviews*. Hoboken, NJ, USA: Wiley, 2008, pp. 239–276, doi: 10.1002/9780470380147.ch4.
- [7] Y.-S. Song, S.-H. Lee, J.-A. Jo, S.-H. Choi, D.-J. Seo, Y.-K. Lee, U. Yang, and W.-J. Jung, "Changes in activity and isozyme patterns of peroxidase and chitinase in kiwifruit pollen," *J. Appl. Botany Food Qual.*, vol. 92, pp. 313–319, Jan. 2019.
- [8] A. Dafni and D. Firmage, "Pollen viability and longevity: Practical, ecological and evolutionary implications," *Plant Systematics Evol.*, vol. 222, nos. 1–4, pp. 113–132, 2000.
- [9] U. Yang, S. G. Wi, H. Kim, J.-H. Moon, S. Oh, M. S. Kim, and S.-H. Lee, "Evaluation of pear pollen germination through colorimetric and cytometric analyses," *Horticultural Sci. Technol.*, vol. 37, no. 4, pp. 427–436, 2019.
- [10] T. Kam Ho, "Random decision forests," in *Proc. 3rd Int. Conf. Document Anal. Recognit.*, vol. 1, Aug. 1995, pp. 278–282.
- [11] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [12] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
- [13] P. Boinee, A. De Angelis, and G. L. Foresti, "Meta random forests," *Int. J. Comput. Intell.*, vol. 2, no. 3, pp. 138–147, 2005.
- [14] W. K. Pratt, *Introduction to Digital Image Processing*. Boca Raton, FL, USA: CRC Press, 2013.
- [15] G. V. Trunk, "A problem of dimensionality: A simple example," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 3, pp. 306–307, Jul. 1979.
- [16] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [17] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1–9.
- [18] S. Haykin, *Neural Networks: A Comprehensive Foundation*. New York, NY, USA: IEEE Press, 1994.
- [19] V. Seviliano and J. L. Aznarte, "Improving classification of pollen grain images of the POLEN23E dataset through three different applications of deep learning convolutional neural networks," *PLoS ONE*, vol. 13, no. 9, Sep. 2018, Art. no. e0201807.
- [20] R. Gallardo-Caballero, C. J. García-Orellana, A. García-Manso, H. M. González-Velasco, R. Tormo-Molina, and M. Macías-Macías, "Precise pollen grain detection in bright field microscopy using deep learning techniques," *Sensors*, vol. 19, no. 16, p. 3583, Aug. 2019.
- [21] R. Navares and J. L. Aznarte, "Geographical imputation of missing poaceae pollen data via convolutional neural networks," *Atmosphere*, vol. 10, no. 11, p. 717, Nov. 2019.
- [22] S. Dunker, E. Motivans, D. Rakosy, D. Boho, P. Mäder, T. Hornick, and T. M. Knight, "Pollen analysis using multispectral imaging flow cytometry and deep learning," *New Phytologist*, vol. 229, no. 1, pp. 593–606, Jan. 2021.
- [23] V. Seviliano, K. Holt, and J. L. Aznarte, "Precise automatic classification of 46 different pollen types with convolutional neural networks," *PLoS ONE*, vol. 15, no. 6, Jun. 2020, Art. no. e0229751.
- [24] N. Khanzhina, E. Putin, A. Filchenkov, and E. Zamyatina, "Pollen grain recognition using convolutional neural network," in *Proc. ESANN*, 2018, pp. 1–6.
- [25] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.
- [26] G. Casalicchio, C. Molnar, and B. Bischl, "Visualizing the feature importance for black box models," in *Machine Learning and Knowledge Discovery in Databases*. Cham, Switzerland: Springer, 2019, pp. 655–670, doi: 10.1007/978-3-030-10925-7\_40.
- [27] Y. Bengio, *Learning Deep Architectures for AI*. New York, NY, USA: Now, 2009.
- [28] P. Hall, N. Gill, M. Kurka, and W. Phan. (2017). *Machine Learning Interpretability With H2O Driverless AI*. [Online]. Available: <http://docs.h2o.ai/driverless-ai/latest-stable/docs/booklets/MLIBooklet.pdf>

[29] M. T. Ribeiro, S. Singh, and C. Guestrin, ““Why should i trust you?": Explaining the predictions of any classifier,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 1135–1144.



**UNG YANG** received the B.S., M.S., and Ph.D. degrees in biological sciences from Chonnam National University, Gwangju, South Korea, in 2004, 2006, and 2013, respectively. He has been working as a Research Professor with the Asian Pear Research Institute, Chonnam National University, since 2018. His research interests include analysis of pollen germination images using machine learning, evaluation of pollen germination through flow cytometry, and exploration

of germination factors affecting pollen performance.



**SEUNGWON OH** received the B.S. and M.S. degrees in statistics from Chonnam National University, South Korea, in 2016 and 2018, respectively. He is currently pursuing the Ph.D. degree with Chonnam National University. His research interests include machine learning algorithm and artificial intelligence methodologies.



**SEUNG GON WI** received the B.S., M.S., and Ph.D. degrees in wood science and landscape architecture from Chonnam National University, Gwangju, South Korea, in 1995, 1999, and 2003, respectively. From 2004 to 2006, he was a Postdoctoral Fellow with the Advanced Radiation Technology Institute, Korea Atomic Energy Research Institute (KAERI). Since 2017, he has been working as a Research Professor with the Asian Pear Research Institute, Chonnam National University,

and is currently participating in interdisciplinary research, especially focused on Korean Pear Export Research. His research interests include pollen morphology and plant anatomy, fruit safety management control, and post-harvest quality control of fruits.



**BOK-RYE LEE** received the Ph.D. degree from Chonnam National University, South Korea, with a focus on physiological changes of metabolic pathway, stress-induced protein, and biochemical in white clover under drought stress. During the Ph.D. degree, she joined Prof. Avicé’s Lab, Université de Caen Normandie, Caen, France, for one year. She worked as a Postdoctoral Researcher with the John Innes Centre, Norwich, U.K., from 2008 to 2010, and with Michigan State University, MI, USA, from 2010 to 2012. Afterwards, she joined as a Research Professor with the Biotechnology Research Institute, from 2012 to 2018. Since 2018, she has been with the Asian Pear Research Institute, Chonnam National University. Her research interests include drought stress responses, hormonal crosstalk, redox-signaling, autophagy, plant-pathogen interaction, plant immunity, and fruit growth modelling.



**SANG-HYUN LEE** received the B.S., M.S., and Ph.D. degrees in horticultural sciences with Chonnam National University, Gwangju, South Korea, in 1998, 2000, and 2003, respectively. He experienced as a Postdoctoral Fellow with Tsukuba University, Japan, and with the Rural Development Administration, South Korea. He joined the Faculty of the Department of Horticulture, Chonnam National University, in 2016. He has been serving as the Director for the Asian Pear Research Institute, since 2017. His research interests include improvement pollen quality and post-harvest physiology in pear fruit.



**MIN-SOO KIM** received the B.S., M.S., and Ph.D. degrees in computer science and statistics with Chonnam National University, Gwangju, South Korea, in 1994, 1996, and 2000, respectively. From 2003 to 2005, he was a Postdoctoral Fellow with the Korea Advanced Institute of Science and Technology (KAIST). In 2007, he joined the Faculty of the Department of Statistics, Chonnam National University. He is currently serving as the Chairman for the Big Data Center, Chonnam National University, since 2015. His research interests include pattern recognition, wavelet, artificial intelligence, and multivariate statistics and analysis of big data according to climate changes.

...